BEAR CREEK SURFACE WATER SIMULATION MODELING DEMONSTRATION

Special Report 13

Kevin L. Anderson

GEOGRAPHY DEPARTMENT

UNIVERSITY OF MINNESOTA

WATER RESOURCES RESEARCH CENTER
UNIVERSITY OF MINNESOTA
ST. PAUL, MN 55108
JUNE 1987

BEAR CREEK SURFACE WATER SIMULATION MODELING DEMONSTRATION

Kevin L. Anderson

ABSTRACT

The role people play in changing the character of the Earth's surface has a profound impact on water resources. Reasonable predictions of the results of human actions would be of enormous benefit to planners. Computer simulations can be used to provide these predictions. As with any analysis technique, these simulations require specific input data and make assumptions about reality that limit their application to watersheds of specific sizes and geographic locations.

This report summarizes a test of the interface between existing computer models and water-resource data in Minnesota's current geographic information system. The study area for this demonstration is the Bear Creek basin of Olmsted County, eighty square miles of rolling farmland near Rochester. We used two "off-the-shelf" computer programs, AGNPS and USDAHL, to estimate runoff for the Bear Creek basin. AGNPS is a cell-based, distributed model that uses the Soil Conservation Service's "Curve Number" method with the Universal Soil Loss Equation to estimate surface runoff and soil erosion for storm events. USDAHL is a non-cellular, fitted model that is calibrated to a basin to estimate runoff over a continuous period of time.

Results from the AGNPS and USDAHL models lead us to several conclusions:

- our current GIS lacks good data on antecedent basin conditions, particularly soil moisture, which is necessary information for storm-event models such as AGNPS;
- a point-count (inventory) approach to data collection is necessary to identify the range of basin characteristics. Describing each area in terms of a single soil, land-cover type, or slope (the current practice in the state's GIS) produces poor results. For example, soils or land covers that occupy only a small fraction of a data cell can still produce the majority of the runoff; and
- the cellular approach to watershed subdivision captures the diversity of hydrologic responses and lends itself to use with a GIS better than a polygon approach.

At present, the state has no model that incorporates all of these features at a scale appropriate to analysis of medium-to-large-sized watersheds. It is this very scale in Minnesota that can benefit from policy decisions and planning based on the use of simulation with a GIS.

BEAR CREEK SURFACE WATER SIMULATION MODELING DEMONSTRATION

Kevin L. Anderson

INTRODUCTION

The role people play in changing the character of the Earth's surface has a profound impact on water resources. Reasonable predictions of the results of human actions would be of enormous benefit to planners. Computer simulations can be used to further this understanding. As with any analysis technique, however, these simulations require specific input data and make assumptions about physical processes that limit their application to watersheds of specific sizes and geographic locations.

This report uses two "off-the-shelf" computer programs and existing data sets to demonstrate the applicability of combining a GIS with simulation. The AGNPS and USDAHL hydrologic models are popular ways to study the effects of changes in soil and land cover characteristics on water resources. The recommendations and conclusions that follow this report are based on our experiences in using these models. My study area for this demonstration is the Bear Creek basin of Olmsted County, Minnesota.

Issues

Several fundamental issues must be considered while reading this report, including:

- how a politician, economist, engineer, hydrologist, or lay person will interpret information contained within a water resources GIS or results generated from a simulation model using data from a GIS;
- how the simulation treats geographic and temporal differences in the distribution of major watershed traits that affect runoff (such as land cover, soils, and terrain); and
- how inherent assumptions regarding the linkages (connections) within the physical world identified by a simulation model (and the degree to which each of these are simplified) affect the range and accuracy of estimates derived from an analysis.

All of these issues are significant and too often ignored. Many are so fundamental as to deserve a discussion in their own right (e.g., Gersmehl, Brown, and Anderson 1987). My intention in this report is not to provide a definitive answer on each issue, but to demonstrate their importance with real situations and to provide a few recommendations that should guide future research.

In this paper, I present runoff estimates for storm events in the Bear Creek basin in order to demonstrate the way in which these two computer simulations "predict" the impact of land cover changes. For reasons I will explain later, these values must not be taken for actual runoff responses to establish a policy decision. They instead represent values estimated by simulation models that have been fitted to the basin for discussion purposes only. As with output from any simulation, *the products of these models indicate only the direction and magnitude of response that might happen for specific changes in land cover. These predictions must be substantiated further before they can be applied in making policy decisions.*

The Bear Creek Watershed

My choice of watersheds was severely constrained by one simple fact -- very few small watersheds in Minnesota have gauging stations, and I need a daily record to calibrate a simulation model. Bear Creek is one of only a handful of Minnesota streams that have a discharge record on a daily basis (and even the period of that record is very short). Bear Creek is also one of the few gauged watersheds in Minnesota that can be found entirely within a county that has a modern soil survey. The Olmsted County survey exists both in its standard published form and as a computerized atlas in the state Soil Survey Information System (SSIS). The SSIS package was developed by the Department of Soil Science and Minnesota Experiment Station at the University of Minnesota in cooperation with the State Planning Information Center and the U. S. Soil Conservation Service. At present, only a dozen or so counties across the state have been digitized into this system, primarily those counties with recently published soil surveys.

Bear Creek is one of three creeks (along with Cascade and Silver) that join the South Fork of the Zumbro River within the city limits of Rochester, Minnesota. The basin, roughly eighty (80) square miles in size, is entirely within Olmsted County and drains prime agricultural land to the south and east of Rochester. Located on the edge of the forested and dissected "driftless" area of Southeast Minnesota, the Bear Creek basin falls within several geomorphic regions. The loess-capped Rochester Drift Plain is by far the largest of these regions (Figure 1). Soils in the basin occur primarily in small mapping units. The soils with high runoff potential are concentrated in the creek valley (Figure 2).

The Simulation Models

The Bear Creek basin has been subject to frequent and costly flooding in the past. The watershed could benefit from prior analysis of future changes using computer models such as the ones discussed here. For example, cost-benefit analyses developed with simulation can minimize cost overruns incurred from unplanned problems encountered after an erosion-control project has started. Several simulation programs developed just for this purpose exist today. This paper looks at two such models, each representing a different approach to hydrologic simulation:

(1) **AGNPS**, the Agricultural Non-Point Source Pollution Model developed locally by the Agricultural Research Service (ARS) office in Morris, the Minnesota Pollution Control Agency (MPCA), and the Soil Science Department at the University of Minnesota (Young et al. 1985); and

(2) **USDAHL**, a soil erosion and runoff prediction model developed by the U.S. Department of Agriculture (USDA) Hydrologic Laboratory. USDA HL-77 is the most recent version available (Holtan and Yaramanoglu 1977) and incorporates several improvements over features found in the initial programs (Holtan and Lopez 1971; Holtan et al. 1975).

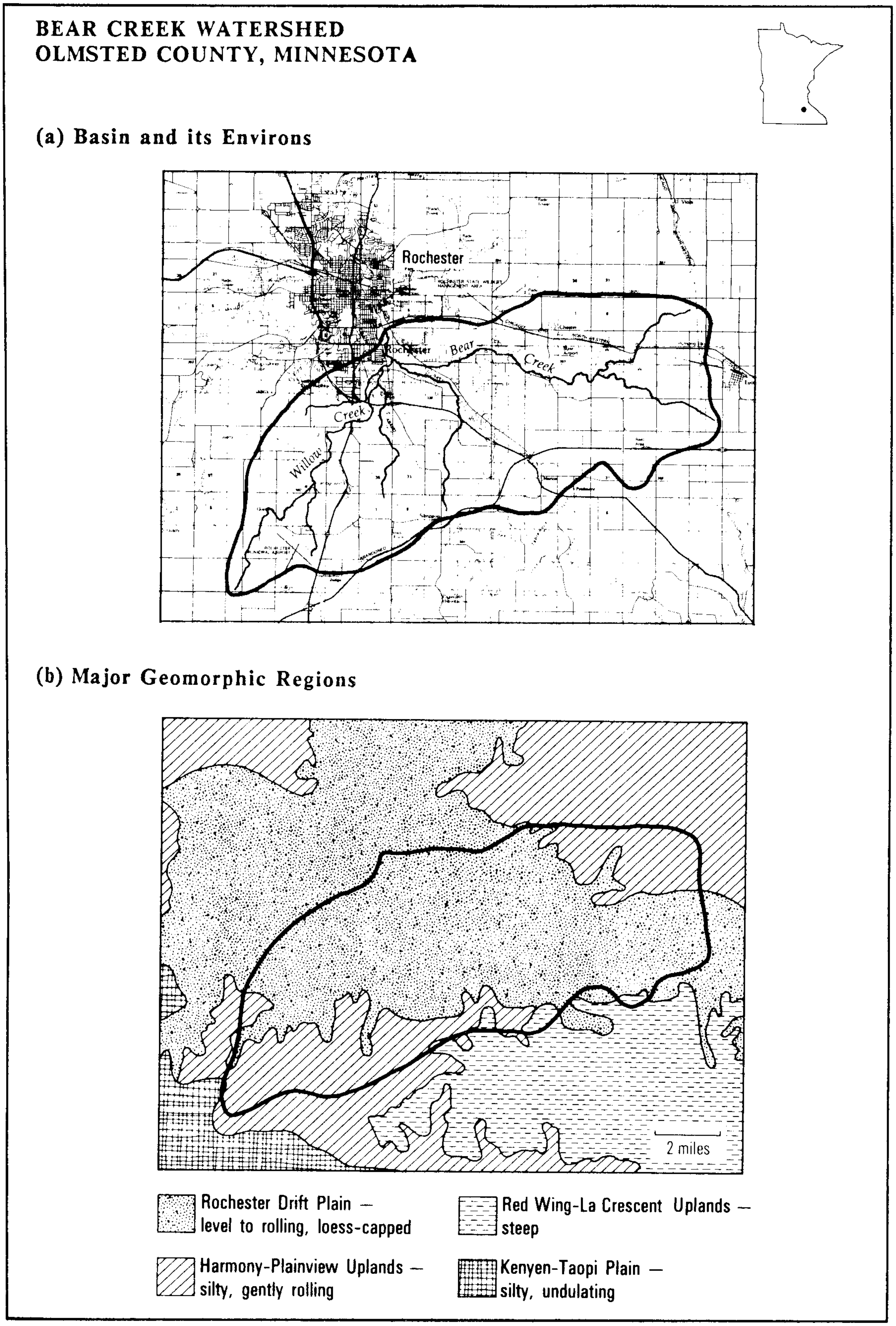
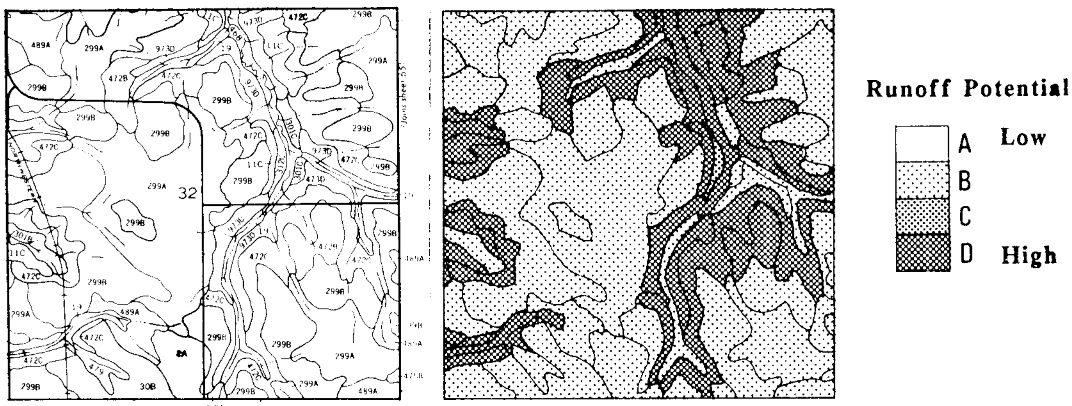

BEAR CREEK WATERSHED
OLMSTED COUNTY, MINNESOTA
(a) Basin and its Environs
Rochester
Bear Creek
Willow Creek
(b) Major Geomorphic Regions
2 miles
Rochester Drift Plain — level to rolling, loess-capped
Red Wing-La Crescent Uplands — steep
Harmony-Plainview Uplands — silty, gently rolling
Kenyen-Taopi Plain — silty, undulating

FIGURE 1

FIGURE 2 -- TYPICAL SOIL LANDSCAPE, OLMSTED COUNTY, MINNESOTA (Section 32, Marion Township)



Soil Mapping Units in Detailed Survey

Soils Classified by Runoff Potential

Both computer programs were originally developed to run on mainframe computer systems, but are now available in versions for today's 16-bit personal computers (such as the IBM PC and its compatibles). All results generated for this report were produced using the microcomputer versions of each model.

The AGNPS model builds upon the CREAMS (Chemicals, Runoff, and Erosion from Agricultural Management Systems) simulation, an earlier model developed for studying small agricultural fields (Knisel 1980). Both rely on the Universal Soil Loss Equation (USLE) to predict soil erosion caused by runoff for the watershed (Wischmeier and Smith 1978).

The model *is a cell-based, distributed simulation*, meaning it subdivides watersheds into a series of smaller areas based on square *cells*, such as 40-acre parcels, each with its own X-Y coordinate, input data, and runoff calculation. The simulation requires data on 20 different parameters for each cell, including:

- the Soil Conservation Service "Curve Number" for the combination of soil and land management in the cell (Soil Conservation Service 1972, 1975 and 1986),
- the USLE parameters on soil erodibility (K), cropping (C) and management practices (P) for each area,
- the slope, direction, and shape of the land surface,
- an average field slope length based on estimates for regions of Minnesota,
- the slope of defined stream channels and a "roughness" value for their bottom,

- a surface condition constant, and
- the coordinate of the cell that receives runoff from this cell.

As an agricultural pollution model, AGNPS also requires information on the application of fertilizer and the size and location of feedlots.

USDAHL, in contrast, does not use square cells to define its space, but instead expects information for two to four *zones* within a watershed. Each zone represents a physically-defined area such as a farmland plateau or river floodplain that contributes runoff and soil sediments to the next zone downhill. Input parameters for zone identification include values for average slope gradient and length, soil depth, infiltration rate, and available water holding capacity. Up to nine different combinations of land cover, crop type, and soil are defined for the basin, with their coverage recorded as the percent of area in each zone. Each crop category, in turn, requires information on its average soil temperature, evapotranspiration rates, rooting depth, and tillage patterns.

AGNPS *is a storm event model*--it estimates runoff and soil erosion for particular storm events, such as the 10-year, 24-hour rainfall at Rochester. Program output from the model, which includes values for sediment movement and peak runoff at the basin outlet and for cells within the basin, is based entirely on a statistical estimate that cannot be fully substantiated given a relatively short historic record of precipitation and runoff. The simulation is designed for small watersheds (less than 30 square miles) and assumes that all water must leave the basin within a 24-hour period. My use of AGNPS stretches this scale restriction considerably. My goal, however, is not to use the output from the model in any way except to test the concept of using simulation models with a GIS for water-resource assessment. To attempt such a test, I needed good soils data and daily runoff measurements, neither which exist for many watersheds between 20 and 30 square miles.

USDAHL *is considered to be a continuous-synthesis model* in that it estimates runoff for a continuous record of rainfall using daily precipitation values for the period in question. Unlike AGNPS, which would need information on watershed conditions prior to the storm event in order to properly control for soil moisture wetness, USDAHL can be "self-calibrating" in its estimates of antecedent moisture conditions. To calibrate itself, the model must be run for a brief period prior to the period of analysis. The dryer a soil is before a storm, the more rainfall that will infiltrate and the less that will runoff. Continuous models are also better suited for including the *base flow* of a stream (the amount of year-round flow usually contributed by groundwater or from water stored in the soil) in its estimates of runoff. It is the base flow that will be visible in a stream channel between the storm events. The USDAHL model, however, has no groundwater component and therefore misses the contribution of groundwater to the recession curves and the dry-weather streamflow. (The user must spend considerable time in deriving good values for USDAHL recession parameters; error in these values greatly reduces the usefulness of the simulation.) USDAHL *is also considered to be a fitted model* in that few of the input parameters have a physical basis to explain the choice of values. The best simulation results come after the model has been calibrated to a known streamflow record. Since the model is fitted to a specific time period, there is little utility in predicting results for changes in parameters outside the known record of each input variable.

The data required by both computer programs can be obtained from precipitation records, aerial photographs, topographic maps, and detailed soil surveys. Depending upon how much data is already available, anywhere from one day to several weeks is necessary to code data for an area the size of Bear Creek. Results can be obtained in only an hour, however, once the data are coded. Although the AGNPS model is able to take information directly from Planning Information Center (PIC) data files, I found that considerable time was still necessary to interpret their data into a meaningful form. (See Gersmehl, Corbett, and Greene 1987 and Corbett and Gersmehl 1987 for an evaluation of PIC data.)

DATA FOR THE BEAR CREEK BASIN TEST

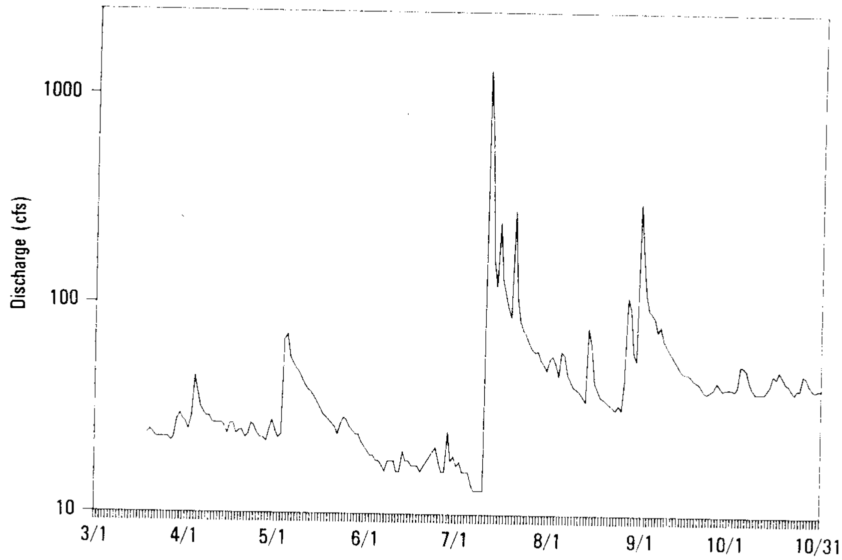
My test of both simulation models involved collecting data by square-mile Public Land Survey sections within the Bear Creek basin for 1981 and comparing predicted runoff against actual streamflow measurements gathered by the U. S. Geological Survey. I used a special USGS streamflow record for Bear Creek, which contains daily discharge measurements from 1981 used to help calibrate a new gauge downstream on the South Fork of the Zumbro River. (Prior to 1981, only peak discharge rates from selected storm events had been measured for 13 years.) This water-discharge record for Bear Creek measured the average daily runoff from March 19th through October 31st in 1981 (Figure 3). Analysis of the record shows that peak flows for 1981 occurred between the 10th and 12th of July (with a maximum recorded discharge of 3,240 ft^3/s on the 11th) from a rainfall between 5 and 7 1/2 inches over the basin. Base flow for Bear Creek appears to be between 20 and 40 ft^3/s, with higher values recorded in the fall months.

Daily precipitation and storm event values used in the Bear Creek simulation were obtained from two stations: from the Rochester Airport south of the city near the middle of the basin, and from the town of Elgin to the northeast. Two versions of a continuous rainfall record were tested with the USDAHL model, one that included just the data from Rochester, and a second that calculated a basin average from values of both rain gages weighted according to the area effectively represented by each. Figure 4 shows the combined daily rainfall record used for the 1981 simulation. These weights (estimated by finding the area in the basin that was closest to each precipitation gage) were 0.7 for the Rochester Airport station and 0.3 for Elgin. (See Drake and Skaggs 1987 for alternative methods of determining values for basin weights.

The goal of any weighted rainfall adjustment is to represent the actual track of a storm through an area. Most summer storms in Minnesota usually produce a spotty distribution of rain that cannot be measured adequately by a single rain gauge. Even though the Rochester Airport lies near the center of the basin, the area is large enough from east to west to allow several storms go through unmeasured. Initially I felt that a single basin average based on proximity to the nearest rain gauges was sufficient to capture that diversity given the size of the basin and the current distribution of gauges. Swerman and Baker (1987) suggest that a minimum of 5 gauges would be necessary to measure total rainfall of individual storms to within five (5) percent of their amount ninety (90) percent of the time.

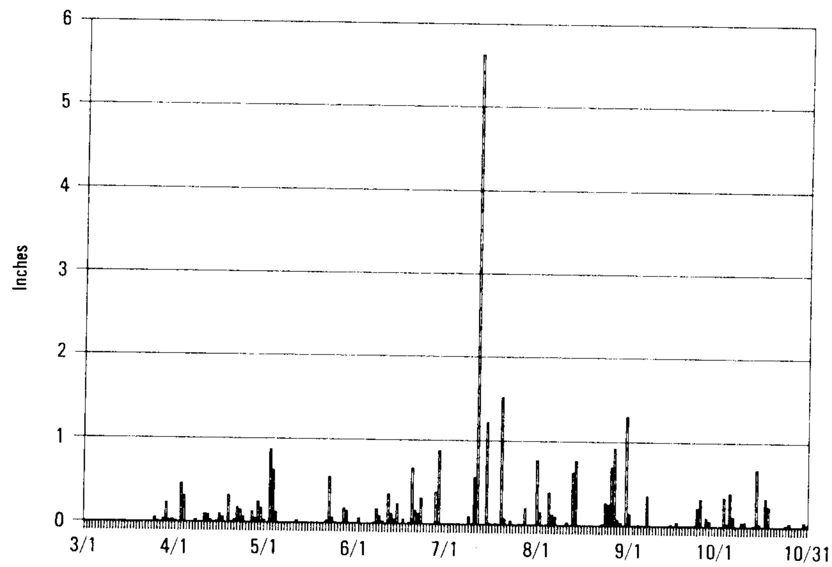
The effect of using a weighted average is considerable: rainfall for the major July storm was 7.47 inches as measured at the airport and only 1.32 inches at Elgin. The combined basin average is 5.6 inches for that particular storm. If only the Rochester Airport measurement is used in a storm event model such as AGNPS, the model overestimates the runoff, whereas use of just the Elgin record would horribly underestimate the storm's effect. A straight average of the two records

FIGURE 3 -- DAILY STREAMFLOW OF BEAR CREEK, 1981



The vertical axis on this graph is based on a *logarithmic* scale, which visually emphasizes the lower base flow values at the expense of under-representing the more pronounced peak flows.

FIGURE 4 -- ESTIMATED DAILY RAINFALL: BEAR CREEK WATERSHED, 1981, USING A WEIGHTED SPATIAL AVERAGE



These rainfall estimates represent weighted averages over the Bear Creek basin, where 70 percent of the precipitation was based on daily measurements taken at the Rochester Airport near the center of the basin, and 30 percent was based on measurements from Elgin just to the northeast of the basin.

(4.4 inches estimated for the July 11th storm) still underestimates the actual rainfall amounts by failing to consider the size of the basin, and supports Swerman and Baker's (1987) conclusion that more gauges are needed.

Eighty-eight square-mile sections formed the "cells" that define the Bear Creek watershed for simulation by the AGNPS computer program. These same eighty-eight sections were grouped into two "zones" for use with the USDAHL model (Figure 5). I used generalized soils and land use information from PIC data files based on 40-acre parcels of land (the MLMIS40 files arranged by counties). The PIC soils data came from *Soil Atlas* maps produced by the Minnesota Experiment Station and Soil Science Department at the University of Minnesota. These maps categorize soil according to drainage class, color, and surface and subsurface texture. The scale of these maps (1:250,000 or four miles to an inch) limits their ability to show fine details. Square miles are the smallest mappable area, and therefore the maps are suitable only for analysis of areas at least a township (36 square miles) in size (Rust et al. 1976; Rust 1986).

In addition to using the PIC files, I collected soil data from the detailed soil survey for Olmsted County (Poch 1980) to help in the interpretation of the Soil Atlas data. Even though the Olmsted County soil survey is part of the new state computerized survey, output options from that GIS proved to be too limiting for my needs. Topographic information, including slope values for land surfaces and stream channels, were derived from large-scale 1:24,000 (approximately 1/3 of mile to an inch) topographic maps. To supplement the MLMIS40 land use data (which was originally coded in 1968 and 1969), I used data from a land-cover inventory derived from newer photography and subdivided each 40-acre parcel by fifths for measurement. I chose not to use a newly-digitized land-use map developed by Olmsted County because the categories did not lend themselves to a hydrologic classification of land cover (See P. Gersmehl, Anderson, Greene, Dunning, C. Gersmehl, and Brown 1987 for a discussion of the hydrologic consequences of a land-use classification).

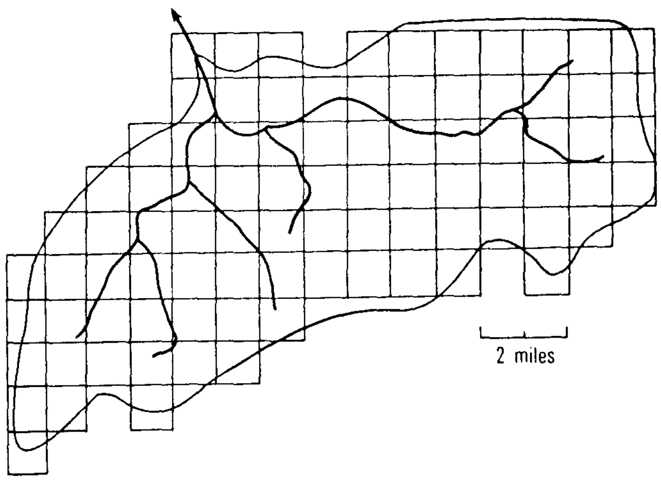
SIMULATION USING AGNPS

Although AGNPS generally overestimated direct-runoff from the basin for every storm event simulated with the model, the overprediction behaved as expected in this demonstration. The simulation calculated that 2.5 inches of rainfall would produce 1.2 inches of runoff out of the basin with a peak discharge of 6,900 ft^3/s (about double the 1981 peak). Similarly, 6.1 inches of rainfall was estimated to produce 4.3 inches of runoff out of the basin with a peak discharge of 25,000 ft^3/s. This flow, although not measured for the July 11th storm of similar magnitude, is not impossible. The maximum peak flow measured by the USGS in the 13 years prior to 1981 was also 25,000 ft^3/s for the Bear Creek. (See the Appendix in this report for a discussion on significant digits in simulation and mapping.)

Although these estimates are of the right order of magnitude, there is uncertainty about these results for several reasons. **First**, square-mile sections are too large to use as "cells" to define a basin for AGNPS. The mathematical equations used to calculate runoff in AGNPS cannot produce realistic results for such large areas. **Second**, the model assumes that all runoff must leave the basin within 24-hours of a storm event. Except in highly urbanized areas, not all runoff can leave an 80-square mile basin in Minnesota in that short a time. This is true even for a 36 square-mile area divided into 40-acre parcels (the maximum size

TWO POSSIBLE WAYS (OUT OF SEVERAL) TO SUBDIVIDE THE BEAR CREEK WATERSHED FOR SIMULATION

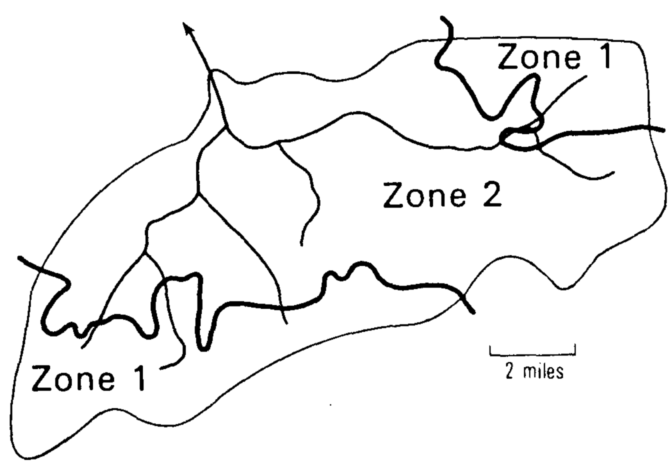
(a) Square-mile grid cells for AGNPS



Square-mile sections from the Public Land Survey system are a compromise between a need for detail (which would favor a small cell of 40 acres or less) and the large size of the basin. As with any cell-based GIS, border cells always present a problem in deciding which areas are to be included and which are to be left out. In Bear Creek, these border ambiquities affect mostly the near-level upland areas. In cell-based simulation, every cell has a code to identify another cell that receives runoff as it is *routed* downhill.

(b) Irregularly shaped zones for USDAHL

Zones are normally chosen to identify major soil and terrain differences between upland, shoulder, slope, and bottomland areas. Zone 1 on this map represents the silty, gently rolling region known as the Harmony-Plainview Uplands; Zone 2 is defined by the loess-capped Rochester Drift Plain. Irregularly-shaped zones (in contrast to square cells) remove ambiquities in borders, but make areal measurements much harder. Runoff is routed downhill from zone to zone based on estimates for the *time to outlet* from each zone.



DEPT OF GEOGRAPHY, UNIV OF MINNESOTA

FIGURE 5

defined for AGNPS). **Third**, the collection of data needed for the AGNPS model is problematic and will affect the accuracy of simulation results. The requirement of just a single data value assigned to each parameter in a "cell" fails to recognize the diversity in hydrologic responses.

For the area around Rochester (and, indeed, in much of the rest of Minnesota), a square-mile section of land is large enough to contain several significantly different combinations of land cover and soil types, each with its own distinct infiltration rate, management practice, and land slope. This problem is significant even for 40-acre tracts of land in our glacial-formed landscape, with numerous hills, marshes, and small lakes providing the often-billed recreational enjoyment and diversity we cherish in the state. Averaging the parameter values (or concentrating on the most extensive characteristics of each mile) ignores the range of runoff responses possible for a storm event.

This geographic diversity is troublesome in the coding of SCS Curve Numbers (as an example of just one input parameter for the AGNPS model). A *curve number* is defined as a value from 25 and 100 that communicates the magnitude of runoff expected from each land cover and soil combination. Higher values suggest more runoff. A curve number of 91 (poor-condition corn on poorly-drained soil) represents approximately 3 inches of runoff from a 4-inch rainfall. A curve number of 51 (good condition alfalfa on well-drained soil) represents only 0.5 inches from the same amount of rain. In the Bear Creek simulation data, many of the sections contain areas with curve numbers that range in value from 55 to 98. For any given cell, therefore, the actual estimate of runoff can vary by nearly a factor of 10 over that predicted by just a single curve number. This range is significant. Research has shown that accurate curve number determinations are necessary for good runoff estimation using the SCS approach employed by AGNPS (Hawkins 1975; see P. Gersmehl, Anderson, Greene, Dunning, C. Gersmehl, and Brown 1987 for a more detailed discussion on the selection of SCS curve numbers).

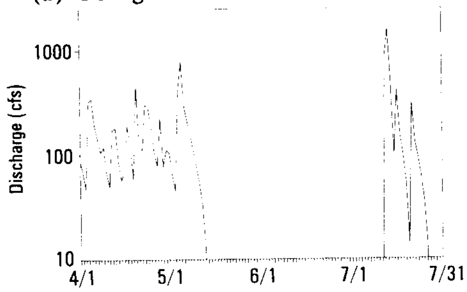
A cell-based simulation approach such as AGNPS can be improved in a number of ways. The most significant improvement is to allow more than one value for a given parameter in each cell. A frequency *count* (inventory) of several curve numbers in each mile will better capture the diversity of land covers. The count approach allows larger cells to be used. A second major improvement is to let each cell's hydrology dictate the rate of runoff from the basin. Do not apply arbitrary time limits (e.g., 24 hours) to runoff, regardless of basin size. Such a limit unnecessarily restricts a model's applicability.

SIMULATION USING USDAHL

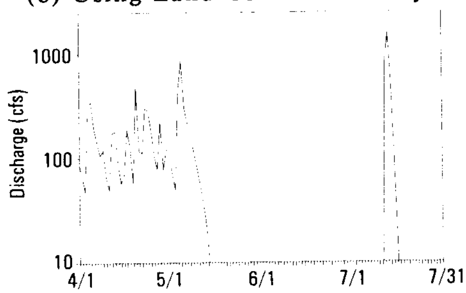
Figures 6a and 6b show the Bear Creek runoff estimates by USDAHL on the basis of data from the two separate land use files (MLMIS 40-acre land-use data and the land-cover inventory based on fifths of a parcel). Both simulations overpredict the runoff flows earlier in the year, most likely due to inaccurate parameters describing rates of soil moisture percolation or evapotranspiration. Both simulations adequately estimate the peak flow from the July 11th storm, although the runoff recession was too short. USDAHL was not successful in simulating base flow for Bear Creek. Lack of base flow in the results will tend to lessen the total flow estimated by the simulation. Incorrect estimates of a discharge hydrograph could also predict a recession time that is too short or a return flow contribution that is too small.

EFFECT OF LAND COVER AND PRECIPITATION CONDITIONS ON SIMULATED RUNOFF FOR BEAR CREEK AS ESTIMATED USING THE USDAHL MODEL

(a) Using 1969 MLMIS Data

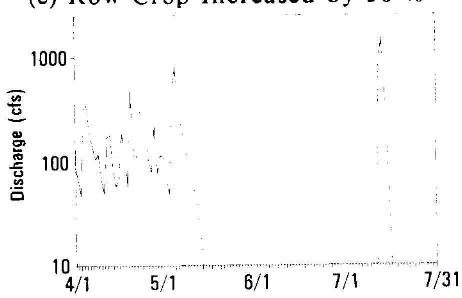


(b) Using Land-Cover Inventory Data

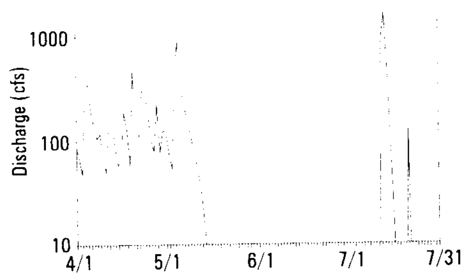


Resolution of both land-use data sets is 40 acres. The 1969 MLMIS study identified a single land-use category for each parcel, whereas the land-cover inventory measures land use in fifths (1/5) of a parcel. Both data sets provide reasonable estimates of runoff in the month of April and for the July 11th storm. Neither simulation was able to estimate base flows between storms events.

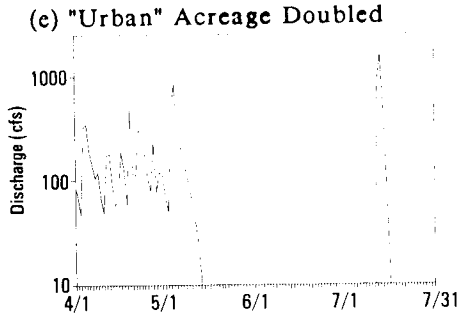
(c) Row Crop Increased by 50 %



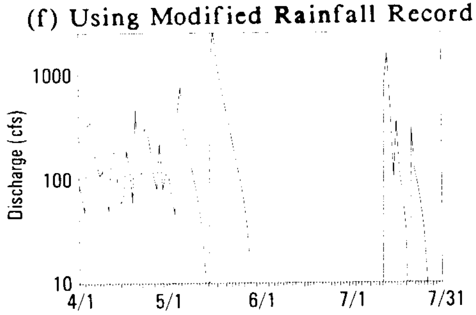
(d) Row Crop Decreased by 50 %



(e) "Urban" Acreage Doubled



(f) Using Modified Rainfall Record



Reasonable modifications to land use in the Bear Creek basin produced little apparent change in simulated runoff. Effects of local land use were "lost" in the simulation, a result of the loose way in which USDAHL locates a particular land use within the basin. Adding a 5.6 inch storm in May to the rainfall record produced nearly twice as much runoff (a peak discharge of 3,000 versus 1,500 cfs) as measured in July, due largely to the near-saturated soil in May.

FIGURE 6

Only two zones (out of four possible zones allowed in the model) were used in these simulations. The zones for Bear Creek were based on differences in soils and topography identified for the major geomorphic regions. The low resolution of available GIS data did not permit definition of more zones. Unlike the arbitrary borders imposed by cells in the AGNPS model, USDAHL zones were allowed to follow irregularly-shaped topographic boundaries. Land cover data, however, was still based on 40-acre parcels.

Simulation models in water resources have their greatest utility in estimating the effects of changes to basin on runoff, not just in representing current conditions. For instance, Figures 6c and 6d present estimated streamflow following a reduction and increase by 50 percent respectively in the amount of cropland within the basin. Similarly, Figure 6e shows the runoff after urban land was doubled to roughly 5,000 acres. It would appear from these tests that the Bear Creek basin makes little response to "minor" land-cover changes (including both small changes over large areas and major changes over small parts of the area) such as the ones chosen here.

As simulated by USDAHL, however, the Bear Creek basin responds much more to the timing of storm events. As an example, USDAHL estimates that the stream discharge in May from 5 1/2 to 7 inches of rain would be roughly 3,000 ft^3/s, or nearly twice the discharge generated later in July from the same size storm (Figure 6f). One should, in general, expect storms of the same size to produce different runoff responses at different times of the year. This higher discharge level is most likely due to a higher antecedent moisture condition in May than in July. Saturated soil conditions will reduce the amount of infiltration. We cannot know for sure unless a number of estimates are generated for different times of the year using a variety of moisture conditions. Even then, it will not be possible to know if these conditions are the cause. *Simulation does not prove causality, it only tests programming linkages.*

As with cell definitions in the AGNPS model, we must be careful with how zones are delimited for USDAHL. Location of particular types of land cover has an affect on the amount and timing of runoff for most basins in Minnesota. For example, the location of urban land within a basin affects the time it takes for runoff to peak after a storm event. When urban land is close to the basin outlet, peak discharge occurs much sooner than if the urban land is at the most distant edge of the basin. Urban land that is distributed throughout a basin should produce the same runoff, but with a lower peak flow, than urban land that is concentrated (Figure 7).

When zones are properly defined for the USDAHL model, they represent distance from the basin outlet and identify differences in the location of various land covers. Zone 1 water flows into Zone 2, Zone 2 runoff flows into Zone 3, and so on. Basin topography (elevation differences, stream network topology) normally dictates how these zones are defined. Not all basins in Minnesota, however, are "normal" with respect to how basic characteristics are distributed up and downstream. Zones for USDAHL, therefore, would not be defined as contiguous and alongside the next zone in downstream order. Even Bear Creek, a relatively simple basin, has that problem. Simulation from these types of areas is tricky; predictions from a model such as USDAHL must be used with caution, and the way various basin characteristics can affect runoff must be understood.

EFFECT OF THE LOCATION OF "URBAN" LAND COVER WITHIN A GENERALIZED WATERSHED ON THE TIMING AND SIZE OF RUNOFF FROM STORM EVENTS

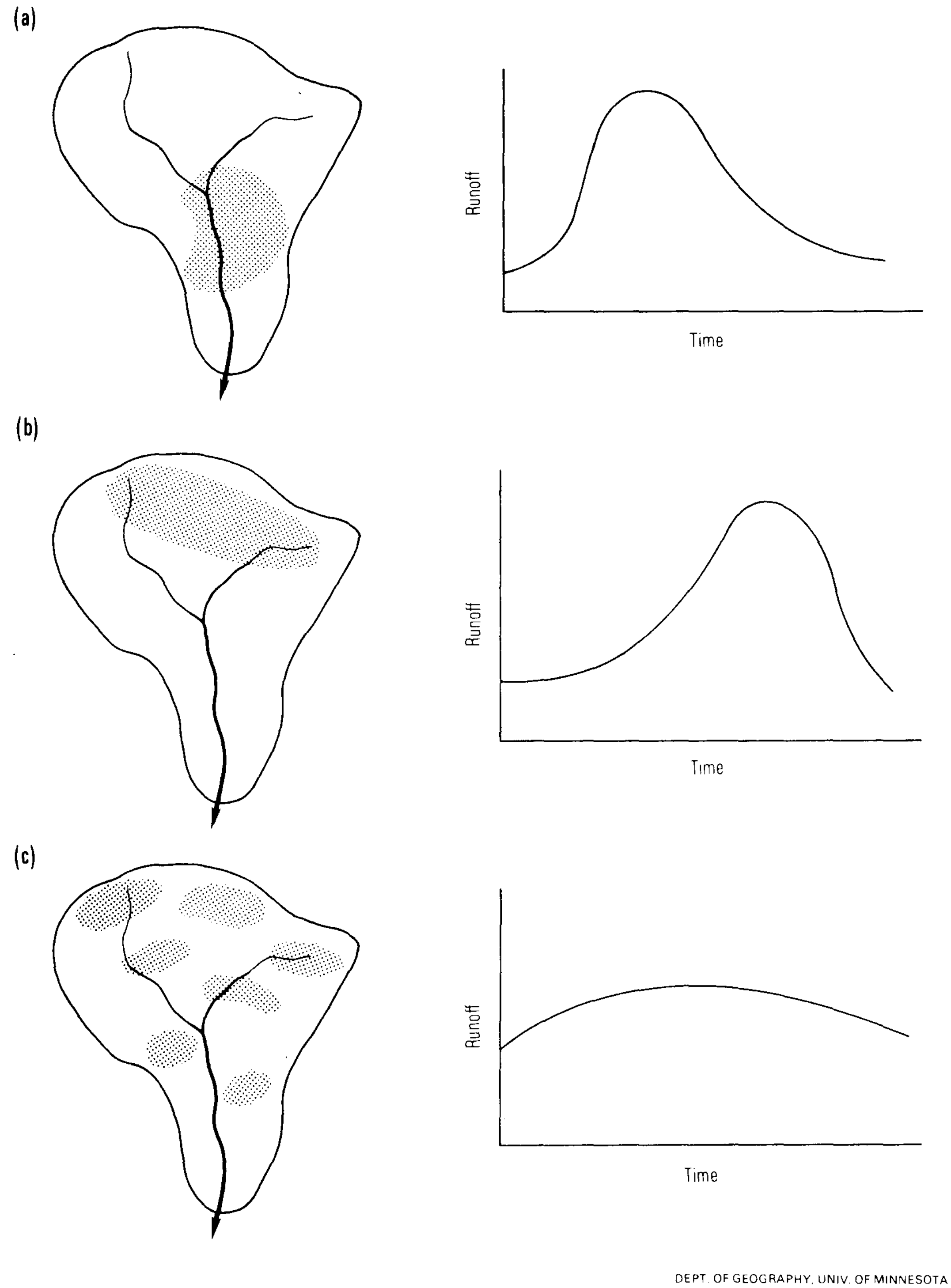


FIGURE 7

RETURN TO THE ISSUES

Several fundamental issues introduced at the beginning of this report permeated the questions posed during our analysis. Each should be discussed further in order to assess how they affect the use of simulation models with a GIS for water-resource assessment. How well these issues stand up to scrutiny will determine the success or failure of a GIS.

1) **The importance of different simulation methods with respect to policy decisions.**

   How a particular simulation subdivides a watershed or time cannot be taken lightly if policy decisions are to be based on their predictions. For instance, a "storm event" approach (used by AGNPS) will tend to underestimate the amount of temporary storage of runoff within a watershed. Many storm-event models assume an unrealistic amount of time (it's too short) for runoff to travel from the most distant edge of a watershed to the basin's outlet. A continuous approach (used by USDAHL) is designed specifically to alleviate these problems by maintaining a record of the basin's recent history and allowing runoff to leave the basin at a more reasonable rate. These assumptions, though related specifically to the analysis process, directly affect the GIS by indicating which data are to be collected for such a study. *We must keep in mind the fact that every simulation program makes assumptions about watershed runoff which do not always match reality.*

2) **The temporal variability of precipitation and land cover.**

   The time during the year when a particular storm takes place affects the amount of runoff that results. For instance, we found that five to seven inches of rainfall in 24 hours (the largest storm in 1981) will elicit different responses from Bear Creek, depending upon whether the storm takes place on July 11 (the actual date) or at some other time in the year. This same rainfall in April would most likely produce much more runoff, because a typical Minnesota April has wetter soil and less ground cover than is normally present later in the summer. A water resources GIS needs to include this temporal component if it is to serve a complex climatic and geomorphic region such as Minnesota. *A watershed is not a constant feature -- it can vary when compared to other basins or to itself at different times of the year.*

3) **The fundamental difference between area-tagging and point-counting procedures.**

   Runoff is the result of several factors, including land cover, soils, terrain, and precipitation. The accuracy of a simulation of runoff depends in part on the accuracy of the input data. A typical soil survey map is an attempt to describe soil for small areas as concisely yet accurately as possible. This is an example of the *tag approach*, in which a map maker tries to draw lines around areas that are reasonably homogeneous. If the watershed being studied is small (e.g., part of a township), the map maker is able to identify many small distinct parcels. To describe a larger watershed or multi-county region, a cartographer must generalize, usually by omitting parcels that are too small to be mapped separately. The inevitable result is an underestimation of the total extent of some kinds of soils.

Oftentimes the omitted soils are precisely the ones that have the most hydrologic significance! (See Anderson et al. 1986; Gersmehl, Brown, and Anderson 1987; Gersmehl, Corbett, and Greene 1987.) The AGNPS model expects only single values for parameters for each subdivision; in effect, it demands a tag approach to mapping.

An alternate procedure uses point sampling within areas to provide a better estimate of the total area covered by particular soil types. This *count method* provides a much more accurate inventory of the total extent of different soils within an area, but it does not claim to show exact location of individual soils. When used with a simulation like USDAHL, the count approach provides a better estimate of the range of hydrologic responses to storm events. Almost every square-mile section in the Bear Creek basin contained a predominance of Group B (lower runoff potential) soils, but most also had a significant area of Group C and D (high runoff potential) soil. Relatively small areas of Group D soil can actually contribute significant amounts of rainfall to streamflow (Anderson et al. 1986). *Thus, an information system cannot classify individual parcels accurately (especially at small scale) and still provide data for a precise estimate of runoff for a large area.* (See Gersmehl, Brown and Anderson 1987 for a more complete description of this issue with respect to soils information.)

CONCLUSIONS AND RECOMMENDATIONS

We cannot design a GIS for all possible hydrologic simulations, nor can we expect a water-resource model to work with all possible GIS, if only for the reason that geographic differences require differences in sensitivity and solutions. Computer models (including AGNPS and USDAHL) represent only one set of tools available to resource planners. As new levels of understanding develop of our physical environment, contemporary models will be replaced by better methods. Design and development of a GIS that can interact with these future tools, yet still meet present objectives, is needed.

Both models used in this demonstration produced credible results for prediction of runoff, with the USDAHL computer model generally estimating stream discharge values for the period of record in 1981 better than the AGNPS program. Both the AGNPS and USDAHL models can use data from a GIS, though with varying degrees of success. They demand different kinds of data -- area-tagging for AGNPS, and point-counting for USDAHL. The simple simulations presented in this report highlight several areas of concern that affect all GIS, but especially those dealing with the natural resources.

Our results from using the AGNPS and USDAHL computer programs to estimate runoff in the Bear Creek basin lead us to these preliminary conclusions regarding the use of simulation in water-resource assessment:

- A storm event model such as AGNPS will not provide adequate results for runoff prediction due to a lack of concern with the timing of rainfall amounts and antecedent basin conditions. Even if basin conditions are part of the input to the simulation, care must be taken to insure that the values used represent reality. When adequate data are available, a continuous model such as USDAHL will better estimate runoff responses, especially if the mathematical equations used in the model have a physical basis in nature.

- A simulation model for runoff prediction needs to use a point-counting (inventory) approach to code a range of characteristics for each subdivision of the watershed. Forcing each area to be identified by a single soil, land-cover type, or slope will produce erroneous results and is a poor basis for policy decisions. At present, no hydrologic simulation exists in Minnesota that uses the inventory approach and the development of one should receive a high priority. The USDAHL models employs a modified form of this inventory approach (by allowing up to nine land cover categories), yet the model is still too limited for most Minnesota landscapes.

- A simulation approach appropriate for Minnesota watersheds must also include the ability to subdivide areas in a fashion that both captures the diversity of hydrologic responses and lends itself to use with a GIS. The cellular approach of AGNPS is useful as long as the size of each unit area can vary to match the total watershed extent. The larger the watershed, generally, the larger the subdivision. AGNPS, as currently designed, is too limited in the size of watershed and cellular subdivision for most regional-scale analyses.

Simulation models paired with a water-resources GIS have utility in planning applications and for the development of policy decisions. I used data from the Bear Creek basin in order to demonstrate how this process works. Even though the models used in this study have problems that limit their applicability in Minnesota, I believe that the simulation process itself has merit. I strongly urge that a better model be developed; in the meantime, the State should take steps to insure that the models and the GIS data are properly used and interpreted within their limitations. With these precautions, they can be effective tools for assessing the impact of environmental changes on water resources.

APPENDIX

DISCUSSION ON THE ROLE OF SIGNIFICANT DIGITS IN SIMULATION AND GIS

For all their benefits, the computer and calculator have been a disservice for thousands, possibly millions, of individuals over the years. And the problem seems to be getting worse. Together, these electronic tools have created the impression that all calculations have results as precise as that indicated on the LED display or printed on paper. Even the cheapest calculator on the market can display at least eight significant digits, plus a decimal point.

There is an old math adage that seems not to be taught to today's student: *you do not report numeric results with any more accuracy than found in the least significant value that went into the calculation.* For lack of a better name, let me call this the **Significant Digits Rule**, or **SDR**. Calculators, and most computer programs, blindly ignore SDR and print as many digits as they possibly can. Students, in turn, dutifully copy every part of the number down for their answer as if it is entirely correct. (Several years of grading homework assignments provides all the proof I need to support this behavior by students.)

In the days before calculators, a deceptively simple, yet effective, twelve-inch device known as a slide rule graced every person's desk. The devices forced the user, even if they did not understand why, to adhere to SDR. Most of the numeric scales found on a slide rule could identify only three, maybe four, digits. Two of these digits were indicated by tic marks on the slide rule; the location of the third digit was estimated. Both intermediate and final results of a lengthy calculation had to be rounded. Scientific notation was used to keep track of the magnitude of the number. No more accuracy was possible unless you worked an equation by hand. In their rush to the computer age, scientists and students quickly grasped the faster methods of calculating, but they left behind some critical understanding along the way.

There are two different ways to interpret SDR. The **first** interpretation is have the the level of accuracy for results determined by the *location* of the least significant digit. For example, if you manipulated a mixture of numbers with digits in the tenth, hundredth, and thousandth place after the decimal, some will argue that your result should be accurate only to tenths. This approach may work for most people, but it breaks down when calculations involve numbers of differing magnitudes. The **second** interpretation is to count the number of digits for each value and base the accuracy on the *shortest* number. If you have a series of numbers with three, four, and five digits identified, the results should be reported with only three significant digits. This interpretation is equivalent in concept to a slide rule's mechanical restriction.

I prefer this second interpretation of SDR. It should be modified, however, by the concept of *implied precision.* Implied precision recognizes the range of digits possible for each individual variable or value in an equation. For instance, rainfall is usually measured in hundredths of an inch in the United States. Possible amounts of rainfall range from 0 to 99.99 inches for most places. A rainfall record, then, has an implied precision of four places. This way, calculations are not penalized by a loss of precision from a rainfall of only .02 inches on one day, and 5.41 inches on another.

The U.S. Geological Survey seems to have accepted, at least in concept, notions behind SDR in their publication of stream-discharge records. For values of daily mean flow, three digits seems to be the norm. Discharge estimates over 1000 ft^3/s have one or more zeros as least significant digits. I would guess this is either a hold-over from pre-computer days or a direct recognition of fairly low levels of accuracy possible from stream gauging equipment and runoff-stage level relationships. (The latter is probably the least accurate set of values because a stream channel profile is continually changing from erosion and deposition.) The USGS break from a three-digit rule when publishing monthly and annual means and totals of discharge, and in estimating acre-feet equivalents.

Why worry about SDR? My use of the AGNPS and USDAHL models in this report provided daily reminders of how SDR is ignored by most computer programs. The peak discharge from the Bear Creek watershed for a 6.1 inch rain (the 10-year, 24-hour storm) was estimated by AGNPS at 24,742.9 ft^3/s -- six significant digits! Most numbers that I input as parameters had only two digits. Where did the extra digits come from? For my discussion in this report, I felt it appropriate to round the result to 25,000 ft^3/s. I cannot believe that a computer program can be so accurate when I know that most of my input parameters were unable to capture much of the spatial variability that must exist in an 80 square-mile watershed.

Yet the program put out some rather impressive and professional looking output that tempted me to accept these results. I understand the rules of significant digits, and I can interpret these values for what they really represent. There are a large number of individuals around, however, who do not understand SDR. Two solutions exist to prevent the misinterpretation of numbers: educate the user, or modify the program. I opt for the latter, since I have a rather pessimistic view of how much education is possible. Computer programs designed for simulation should automatically round their output to the appropriate level of significance based on implied accuracy of input parameters. This accuracy is a "fuzzy concept", since it not just the number of significant digits that are entered which must be considered, but also the extent of knowledge or accuracy we can expect during data collection.

The Role of SDR in GIS

Here is where SDR fits into the new age of geographic information systems. The simulations discussed in this report depend heavily upon data that describe spatial characteristics. These data in turn come from maps, field samples, and basic beliefs in how a phenomenon varies over an area. We obviously cannot visit every spot on earth and catalogue all characteristics about the place. Spatial data must always be *generalized*, or reduced in their level of detail. With maps, scale implies a direct relationship between the size of area represented and the amount of data that can be presented. *Large-scale* maps, such as the standard 1:24,000-scale (approximately 3 inches on the map for every mile on the earth's surface) topographic map of the USGS, can provide more detail than is possible on a 1:250,000 *medium-scale* (one inch represents four miles) map covering the same area. The scale of a map dictates the level of generalization inherent in the data; the contrary is also true, since the level of detail in a data set should imply an appropriate scale of presentation.

GIS bring together data from differing sources and scales of generalization into a common relational setting. Each map (or variable) is coded into the GIS according to its appropriate scale. We introduce an additional level of generalization with a cell-based GIS by our choice of cell size. These generalizations are easily understood when analyzing single data sets. Trouble arises, however, when we compare two or more data sets that have differing scales or resolution (which is most likely the case). In a vector-oriented GIS, which records the outlines of individual mapping units, the combination of two maps could easily result in twice as many mapping units, each with a size that is smaller than the areas found on the original maps. This new minimum size will inevitably be smaller than allowed by either source map.

The rule of significant digits applies also to maps. Maps should be displayed in a GIS with a resolution based on the least significant (largest, coarsest, least detailed) mapping unit that went into the creation of the map. Let's call this the **Significant Mapping Unit Rule**, or **SMUR**. For example, the Soil Atlas data discussed in the paper has a minimum mapping unit size of one square mile. No category of soil with less than a square mile of area will be identified on those maps. Any data that is compared to those soil mapping units, say for instance detailed landholding data with mapping units of about 10 acres, must be generalized to square-mile areas, and not the other way around.

Detailed county soil surveys use an interesting level of generalization that is worth discussing with respect to SMUR. A *soil mapping unit*, defined to have a minimum area of 1.5 acres, can have between 10 and 15 percent of its area in a soil different than the one after which the unit is named. These additional soils are known as *inclusions* (Soil Survey Staff 1975, p. 408-409). A soil mapping unit in a detailed survey, then, is a *spatial probability unit*. For any spot you choose to dig a hole, you have at least a 9 in 10 chance or better of identifying the named soil. You have a 1 in 10 chance of being wrong. You can never have 1 in 1 odds; the physical environment is just not set up to be perfect. Even for the named soil being mapped, the characteristics by which we know this soil are actually a continuum centered on a modal concept (or mean characteristics) that gradually change through time into characteristics of other soils.

GIS managers can learn something from the soil scientist. Their soil-mapping-unit approach is very similar to the inventory approach advocated in this paper. No data set can be perfect. The generalizing process introduces errors or inconsistencies that reduce the ability to describe any particular place with confidence. When data sets are combined, each mapping unit becomes a probability statement for identification of the all characteristics in one place. A regional inventory approach to data collection specifically implies that we have only rough estimates of the amount of the named characteristic in an area, but we cannot tell specifically where they are. The odds must be played in locating actual places.

To most geographers, discussion of these matters is of trivial concern. Geographers are trained from day one to understand these implied levels of generalization and resolution with scale changes. It's old hat. When we see a published map, our focus is automatically adjusted to interpret correctly the amount of detail in the message portrayed by the map. Yet, even with this education, in our rush to introduce the computer to cartography and GIS, we've forgotten many of the basic rules.

As with the lack of SDR built into computer simulations, current GIS do not adequately limit use of coded data. I do not know of a single GIS package that records with each data file information on the scale of the original data source, resolution, level of generalization, purpose, projection type, or a number of other basic characteristics that limit use of that data. No cell-based GIS restricts my subdivision of cells into areas smaller than the minimum mapping unit size. I can combine two data sets of differing scales and display (and therefore interpret) the results at the resolution of the most detailed file. Printing information about map limitations only on the printed products that accompany a map from a GIS is not enough insurance against the misuse or misinterpretation of the data.

We need to reaffirm basic guiding principles of the least significant digits and minimum significant mapping unit rules. I'm not so concerned about the educated users; most of the users of simulation and GIS unfortunately are not educated in these matters. GIS managers need to make sure that their product is used correctly, if only to protect against liability concerns.

REFERENCES

Agricultural Experiment Station, University of Minnesota. 1973. *Minnesota Soil Atlas*, St. Paul sheet, Miscellaneous Report 120-1973, 1:250,000-scale map and report, St. Paul, MN: University of Minnesota.

Anderson, Kevin L., J. D. Corbett, C. A. Gersmehl, D. A. Brown, and P. J. Gersmehl. 1986. "Cell size and simulation modeling of water resources at a regional scale," in Bernard J. Nieman, Jr. (editor), *Proceedings*, Urban and Regional Information System Association (URISA), Denver, Volume I, pp. 203-213.

Corbett, John D., and Philip J. Gersmehl. 1987. "Terrain Data for a Water Resources GIS," Chapter 11 in D. A. Brown and P. J. Gersmehl (editors), *File Structure Design and Data Specifications for a Water Resources Geographic Information System*, Water Resources Research Center Special Report No. 10, St. Paul, MN: University of Minnesota.

Drake, J., and R. H. Skaggs. 1987. "Climatic Network Density Analysis," Chapter 4 in D. A. Brown and P. J. Gersmehl (editors), *File Structure Design and Data Specifications for a Water Resources Geographic Information System*, Water Resources Research Center Special Report No. 10, St. Paul, MN: University of Minnesota.

Gersmehl, Philip J., K. L. Anderson, R. P. Greene, N. P. Dunning, C. A. Gersmehl, and D. A. Brown 1987. "Hydrologic Classification of Land Cover," Chapter 6 in D. A. Brown and P. J. Gersmehl (editors), *File Structure Design and Data Specifications for a Water Resources Geographic Information System*, Water Resources Research Center Special Report No. 10, St. Paul, MN: University of Minnesota.

Gersmehl, Philip J., D. A. Brown, and K. L. Anderson. 1987. "File Structure and Cell Size Considerations for a Water Resources GIS," Chapter 2 in D. A. Brown and P. J. Gersmehl (editors), *File Structure Design and Data Specifications for a Water Resources Geographic Information System*, Water Resources Research Center Special Report No. 10, St. Paul, MN: University of Minnesota.

Gersmehl, Philip J., J. D. Corbett, and R. P. Greene. 1987. "Soil Data for a Water Resources GIS," Chapter 10 in D. A. Brown and P. J. Gersmehl (editors), *File Structure Design and Data Specifications for a Water Resources Geographic Information System*, Water Resources Research Center Special Report No. 10, St. Paul, MN: University of Minnesota.

Hawkins, Richard H. 1975. "The importance of accurate curve numbers in the estimation of storm runoff," *Water Resources Bulletin* 11:887-891.

Holtan, H. N., and N. C. Lopez. 1971. *U. S. Department of Agriculture HL-70 model of watershed hydrology*, Technical Bulletin No. 1435, Washington, D.C.: USDA Agriculture Research Service.

Holtan, H. N., G. J. Stittner, W. H. Hanson, and N. C. Lopez. 1975. *USDAHL-74 revised model of watershed hydrology*, Technical Bull. No. 1518, Washington, D.C.: USDA Agriculture Research Service.

Holtan, H. N., and M. Yaramanoglu. 1977. *A user's manual for the Maryland version of the USDA-HL model of watershed hydrology*, Miscellaneous Publication 918, College Park, Md.: University of Maryland, Agriculture Experiment Station.

Knisel, Walter G., editor. 1980. *CREAMS, A field scale model for Chemicals, Runoff, and Erosion From Agricultural Management Systems*, USDA Conservation Research Report No. 26, Volume I, Washington, D.C.: U.S. Government Printing Office.

Poch, George A., Soil Conservation Service, USDA. 1980. *Soil Survey of Olmsted County, Minnesota.* St. Paul, MN: USDA SCS.

Rust, Richard H., University of Minnesota, Department of Soil Science. 1986. Personal communications.

Rust, Richard H., H. R. Finney, L. D. Hanson, and H. E. Wright, Jr. 1976. "High-altitude Photography in the Development of a Generalized Soil Map," *Soil Science Society of America Journal* 40:405-409.

Soil Conservation Service, USDA. n.d. *Hydrology Guide for Minnesota*, St. Paul, MN: Soil Conservation Service.

Soil Conservation Service, USDA. 1972. *National Engineering Handbook*, Section 4, Hydrology, Washington, D.C.: U.S. Government Printing Office.

Soil Conservation Service, USDA. 1986. *Urban hydrology for small watersheds* (Second Edition), Technical Release No. 55, Washington, D.C.: U.S. Government Printing Office.

Soil Survey Staff, Soil Conservation Service. 1975. *Soil Taxonomy*, Agriculture Handbook No. 436, Washington, D.C.: USDA SCS.

Swerman, R., and D. Baker. 1987. "Precipitation Network Density Requirements for Short-Term Analysis, " Chapter 3 in in D. A. Brown and P. J. Gersmehl (editors), *File Structure Design and Data Specifications for a Water Resources Geographic Information System*, Water Resources Research Center Special Report No. 10, St. Paul, MN: University of Minnesota.

U.S. Geological Survey, Water Resources Division. 1982. *Water Resources for Minnesota, Water Year 1981*, Volume 2, Upper Mississippi and Missouri River Basins, St. Paul, MN: USGS.

Wischmeier, W. H., and D. D. Smith. 1978. *Predicting rainfall erosion losses*, USDA Agricultural Handbook No. 537, Washington, D.C.: U.S. Government Printing Office.

Young, Robert A., C. A. Onstad, D. D. Bosch, and W. P. Anderson. 1985. *AGNPS I, Agricultural Nonpoint Source Pollution Model: A Large Watershed Analysis Tool; A Guide to Model Users*, Draft Copy, Morris, MN: Agricultural Research Service.